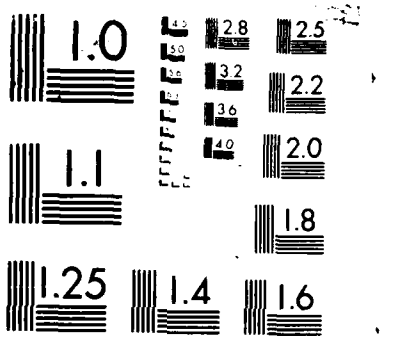
AD-A177 441 ERROR-RESISTANT NARROWBAND VOICE ENCODER(U) NAVAL RESEARCH LAB WASHINGTON DC G S KANG ET AL. 26 DEC 86 NRL-9018 1/1

UNCLASSIFIED F/G 17/2 NL



Naval Research Laboratory

Washington, DC 20375-5000 NRL Report 9018 December 26, 1986

Error-Resistant Narrowband Voice Encoder

G. S. KANG AND W. M. JEWETT

Information Technology Division

Approved for public release; distribution unlimited

A177-14-1

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) NRL Report 9018	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Research Laboratory 6b OFFICE SYMBOL (*If applicable*) Code 7526	7a NAME OF MONITORING ORGANIZATION
6c ADDRESS (*City, State, and ZIP Code*) Washington, DC 20375-5000	7b ADDRESS (*City, State, and ZIP Code*)
8a NAME OF FUNDING / SPONSORING ORGANIZATION (See page ii) 8b OFFICE SYMBOL (*If applicable*) PDE 110	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (*City, State, and ZIP Code*) Arlington, VA 22217 Washington, DC 20360	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: (See page ii) PROJECT NO: (See page ii) TASK NO: WORK UNIT ACCESSION NO: DN 280-209

11 TITLE (*Include Security Classification*)
Error-Resistant Narrowband Voice Encoder

12 PERSONAL AUTHOR(S)
Kang, G.S. and Jewett,* W.M.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (*Year, Month, Day*)	15 PAGE COUNT
Inte	FROM ____ TO ____	1986 December 26	30

16 SUPPLEMENTARY NOTATION
*retired

17 COSATI CODES			18 SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)	
FIELD	GROUP	SUB-GROUP	HF modem	Linear predictive coding
			Bit error protection	Line-spectrum pairs
			Speech intelligibility	800-b/s voice encoder

19 ABSTRACT (*Continue on reverse if necessary and identify by block number*)

One of the major causes of speech quality degradation in digital voice communication is bit errors introduced by the transmission channel. Until now we did not have an effective way to combat bit errors to improve tactical voice communication. For example, a more robust voice terminal that requires a wider transmission bandwidth would not be helpful to tactical communicators because most of them must rely on narrowband channels. Even if some of them may have access to wideband channels, the limited platform space does not allow them to carry both the narrowband voice terminal based on the Government-standard voice algorithm (that would interoperate with all narrowband users) and an additional more robust terminal.

To circumvent this difficulty, we have developed an approach in which voice information is initially encoded at a low data rate (i.e., 800 b/s) and then redundancies for error protection are added to a bit rate that is compatible with transmission over narrowband channels. The necessary software can be integrated into the current narrowband voice terminal so that narrowband communicators have the option of either using the
(Continues)

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL George S. Kang	22b TELEPHONE (*Include Area Code*) (202) 7678-2157 22c OFFICE SYMBOL Code 7526

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

☆U.S. Government Printing Office: 1985—607-047

8a. NAME OF FUNDING/SPONSORING ORGANIZATION

Office of Naval Research
Space and Naval Warfare Systems Command

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.
61153N	RR021-05-42
33904N	X7290-CC

19. ABSTRACT (Continued)

Government-standard 2400-b/s linear predictive coder (LPC) or the more robust voice algorithm. According to our simulation using a high-frequency channel, the new algorithm has a 4 dB advantage over the 2400-b/s system in terms of signal energy to noise density ratio when the bit error rate is 1 to 2%. In essence, the performance of the 800-b/s system is that of the 2400-b/s system with 2.5 times more received signal energy.

CONTENTS

ERROR-RESISTANT NARROWBAND VOICE ENCODER

INTRODUCTION

Tactical voice communications are often brief, but the success of the mission and even the lives of the personnel are often dependent on the reliable transmission of a few critical messages. The linear predictive coder (LPC) operating at 2400 bits per second (b/s) will be widely deployed to support tactical voice communication over narrowband channels. It provides good speech intelligibility in error-free conditions; however the speech quality degrades rather quickly in the presence of bit errors. As indicated in Fig. 1, intelligibility of LPC-processed speech is poor at a bit-error rate of 3%. (Unless otherwise stated, the 2400-b/s LPC referred to is the Government standard 2400-b/s LPC defined by Federal Standard 1015 [1]).

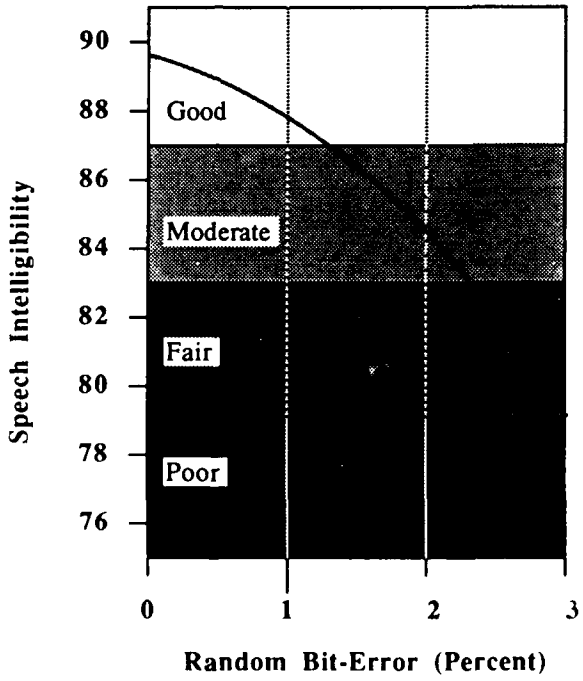


Fig. 1 — Speech intelligibility of the 2400-b/s LPC in terms of random bit-error rate. As noted, speech quality becomes poor if the bit-error rate exceeds 3%. (The descriptors "good," "fair," etc. have been recently adopted by the DoD Digital Voice Processor Consortium.) The lack of robustness is partly caused by the fact that an error in any one of the LPC coefficients alters the speech spectrum over the entire passband. It is interesting to note that earlier channel vocoders were more error-resistant because an error in one channel output affected the speech spectrum only in one narrow frequency band.

Manuscript approved August 20, 1986.

Thus the availability of a more robust voice terminal seems highly desirable. There are, however, two problems in providing an improved capability to *narrowband tactical communicators*:

- The data rate must be low enough to permit transmission over narrowband channels that have a bandwidth of approximately 3 kHz. Most tactical communicators cannot use wideband systems because they do not have access to wideband channels.

- An improved voice processor in a separate package will not help most tactical communicators because their platforms are too congested to carry more than one voice terminal (e.g., amphibious vehicles, high-performance aircraft, armored personnel carriers, jeeps, or tanks). Certainly tactical radio operators operating on foot (Fig. 2) must rely on a single voice terminal, and it must be the 2400-b/s LPC because it is the only narrowband voice processor that has been standardized for interoperability.

The most practical and cost-effective way of providing tactical communicators with improved capability is to integrate the improved voice processing and modem software into an existing 2400-b/s LPC terminal, such as the Advanced Narrowband Digital Voice Terminal (ANDVT). Narrowband users would then have both the 2400-b/s LPC and the improved voice processor without requiring a new radio transmitter, antenna, central processing unit (CPU), packaging, communication security (COMSEC) unit, etc. The operator may manually select one of the two modes. The transmitter and receiver may alternately probe the channel during the preamble period, and the transmitter may select automatically a preferred mode based on the channel conditions. The resulting voice terminal is an example of an *expert system*. As technology advances, the voice terminal could have more elaborate voice processing and error protection algorithms.

To increase the robustness of the narrowband voice processor under conditions of channel bit errors, we have encoded the speech at a low rate (i.e., 800 b/s) and let coding and modulation bring up the data rate so that it is compatible with transmission over the narrowband channel. Note that protection of voice information need not be as sophisticated as protection of digital data because speech has many redundancies and the powerful human brain deciphers the information. According to extensive test data we have collected from various voice processors, LPC-processed speech is intelligible even under 1 or 2% errors. Actually, tactical communication can function under even worse error conditions because:

- tactical communicators use a limited and highly specialized vocabulary consisting of words and phrases that are designed to be easily distinguished in poor signal-to-noise conditions;

- the type of information that is likely to be communicated is also highly dependent on the nature of the mission and the stage in the sequence of the mission so that the communicators know what to expect;

- tactical communicators are accustomed to poor-quality speech, and can accept a less-than-ideal voice terminal if there is no other way to communicate.

Hence, even a slightly improved voice processor would be a help to tactical communicators.

The implementation of a robust narrowband voice terminal presents many technical obstacles. The most difficult problem is to devise a low rate voice processor capable of providing highly intelligible speech. We have been working in this area for nearly a decade, and only recently we have succeeded in implementing what appears to be a satisfactory 800-b/s voice processor [2,3]. In terms of the Diagnostic Rhyme Test (DRT), the intelligibility of the 800-b/s voice processor is only 1.4 points below that of the 2400-b/s LPC. We describe this voice processor in this report.

From pg 41 of "The Government Standard Linear Predictive Coding Algorithm LPC-10," T E Tremain, *Speech Technology*, April 1982, Vol 1, No 2, Copyright 1982, used by permission

Fig. 2 — A tactical radio communicator operating on foot. Similar to tactical radio communicators operating in amphibious vehicles, armored personnel carriers, jeeps and tanks, he cannot carry more than one voice terminal. The purpose of this report is to describe an improved voice processing and modem software that can be incorporated into the 2400-b/s LPC so that the operator can choose either the 2400-b/s LPC or a more error-resistant narrowband voice mode

frames only [1]. Since the unvoiced speech spectrum does not have predominant resonant frequencies, only four LPC coefficients are transmitted for each unvoiced frame. Thus the 21 bits used to encode the fifth through tenth LPC coefficients are freed. By using these 21 bits, the four most significant bits (MSBs) of the amplitude parameters and the first four reflection coefficients are protected (Table 1). Because silence is transmitted as unvoiced frames, the most apparent benefit of this particular error protection is a reduction of loud "pops" during silence periods. This is because the amplitude parameter that controls the loudness of the synthesized speech will have fewer errors.

Table 1 — Error-Protected Unvoiced Speech Data of the 2400-b/s LPC. The bits indicated by shaded blocks are protected by an (8,4) Hamming code. The seven-bit pitch/voicing parameter has some redundancies because there are only 61 possible pitch values. Thus a one-bit error can be decoded correctly in seven zeros that indicate the unvoiced state.

Speech Parameters	MSB						LSB
Pitch/Voicing	1	2	3	4	5	6	7
Amplitude	1	2	3	4	5		
Ref. Coeff. 1	1	2	3	4	5		
Ref. Coeff. 2	1	2	3	4	5		
Ref. Coeff. 3	1	2	3	4	5		
Ref. Coeff. 4	1	2	3	4	5		
Sync	1						

The previously mentioned ANDVT employs more powerful error protection in the high-frequency (HF) modem [8,9]. Among the 54 bits of voice data from each frame (a frame rate of 44.44 Hz), the perceptually most significant 24 bits (Table 2) are error-protected by a Golay (24,12) code. A total of 78 bits is modulated on 39 tones, each separated by 56.25 Hz. The transmission rate is thus 44.444 frames/s with 78 bits/frame for a total of 3466.67 b/s. This additional 1066.67 b/s over 2400 b/s improves the performance significantly as shown in Fig. 16b.

Previous Efforts on Very Low Data Rate Voice Encoding

For many years we have been investigating voice encoders operating at data rates between 600 and 800 b/s (Table 3). Since this is approximately 1% of the data rate of unprocessed digitized speech, some degradation of speech intelligibility is inevitable. Only recently we have been able to devise a voice processor capable of generating high-quality speech at 800 b/s. The Diagnostic Rhyme Test (DRT) for three male speakers over the 800-b/s system is 87.0. Currently, this is the highest score attained by any voice processor operating at a fixed data rate of 800 b/s. The most striking difference between this voice processor and others is the use of new speech parameters called line spectrum pairs (LSPs). We discuss the various aspects of the LSPs from pages 7 through 17.

Table 2 — Speech Data Protected by the ANDVT Modem. The first four reflection coefficients are more critical to the speech spectrum than the remaining reflection coefficients. Likewise a pitch error is readily perceived by the listener. Hence, MSBs of these parameters, indicated by shaded blocks, are protected.

Speech Parameters	MSB						LSB
Pitch/Voicing	1	2	3	4	5	6	7
Amplitude	1	2	3	4	5		
Ref. Coeff. 1	1	2	3	4	5		
Ref. Coeff. 2	1	2	3	4	5		
Ref. Coeff. 3	1	2	3	4	5		
Ref. Coeff. 4	1	2	3	4	5		
Ref. Coeff. 5	1	2	3	4			
Ref. Coeff. 6	1	2	3	4			
Ref. Coeff. 7	1	2	3	4			
Ref. Coeff. 8	1	2	3	4			
Ref. Coeff. 9	1	2	3				
Ref. Coeff. 10	1	2					
Sync	1						

Table 3 — Our Previous Efforts on Low Data Rate Voice Processor Development

Year	Effort	Parameters	Real Time	Data (b/s)	DRT	Ref.
1976	In-house	Formant Frequencies	No	600	79.9 (1M)*	11,12
1980	Contract	Reflection Coefficients	No	800	80.0 (2M)	13
1981	Contract	Reflection Coefficients	Yes	800	78.3 (3M)	14
1983	In-house	Reflection Coefficients	Yes	800	82.8 (3M)	15
1984	Contract	Reflection Coefficients	Yes	800	79.7 (3M)	16
1985	In-house	Line-spectrum Pairs	No	800	87.0 (3M)	2,3

* One male speaker

BLOCK DIAGRAM

In our approach, an improved speech encoder is an extension of the 2400-b/s LPC. The speech parameters are generated by the standard 2400-b/s LPC analyzer. As indicated in the block diagram shown in Fig. 4, the 2400-b/s LPC may be converted to an error-resistant LPC by adding the following three computational modules:

- coefficient converter
- 800-b/s voice encoder
- error protector

These three blocks are discussed in the following sections.

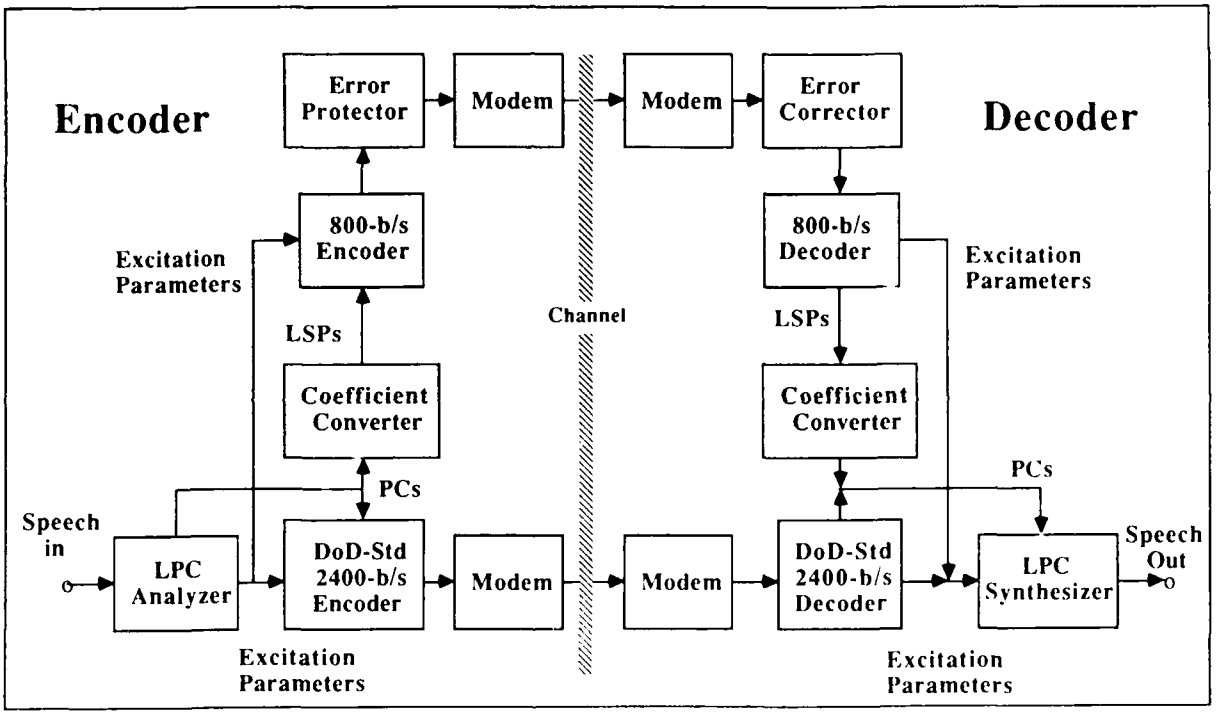


Fig. 4 — Block diagram of dual-mode narrowband voice processor. Addition of the grey blocks converts the 2400-b/s LPC to a more robust voice processor based on 800-b/s LPC (denoted by 800-b/s).

Figure 4 is a block diagram that shows the speech parameters used in the generation of the 800-b/s bit-stream. We recommend the use of unquantized speech parameters rather than a rate-conversion approach that uses quantized parameters since this produces higher speech intelligibility [14].

COEFFICIENT CONVERSION

The improved voice processor converts the set of prediction coefficients (PCs) generated by the LPC analysis into a set of line spectrum pairs (LSPs), and vice versa. We present these conversion algorithms below.

Definition of LSP

The LPC analysis filter converts speech samples to prediction residual samples. Since a residual sample is defined as the difference between the input speech sample and predicted sample (i.e., the sample estimated by a weighted sum of past samples), the transfer function of the LPC analysis filter $A(z)$ may be expressed as

$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \ldots - a_n z^{-n}, \tag{1}$$

where a_n is the n^{th} prediction coefficient. Prediction coefficients are obtained by minimizing the mean-square value of the prediction residual, since the LPC synthesizer is the inverse of the LPC analysis filter, $1/A(z)$. Prediction coefficients are convenient parameters for the LPC analysis/synthesis because they are obtained directly through the LPC analysis. A serious limitation, however, is that an error in one coefficient affects the speech spectrum over the entire passband.

The LPC analysis filter $A(z)$ may also be expressed in the factored form;

$$A(z) = \prod_{i=1}^{n/2} (1 - z_i z^{-1}) \tag{2}$$

where z_i is the ith root of the LPC analysis filter (Fig. 5). The advantage of encoding roots is that an error in one root affects the speech spectrum near that frequency. The roots of the LPC analysis filter have never been used as filter parameters because a fixed-point arithmetic unit (often used in the 2400-b/s LPC) cannot successfully extract these roots from a 10th order polynomial.

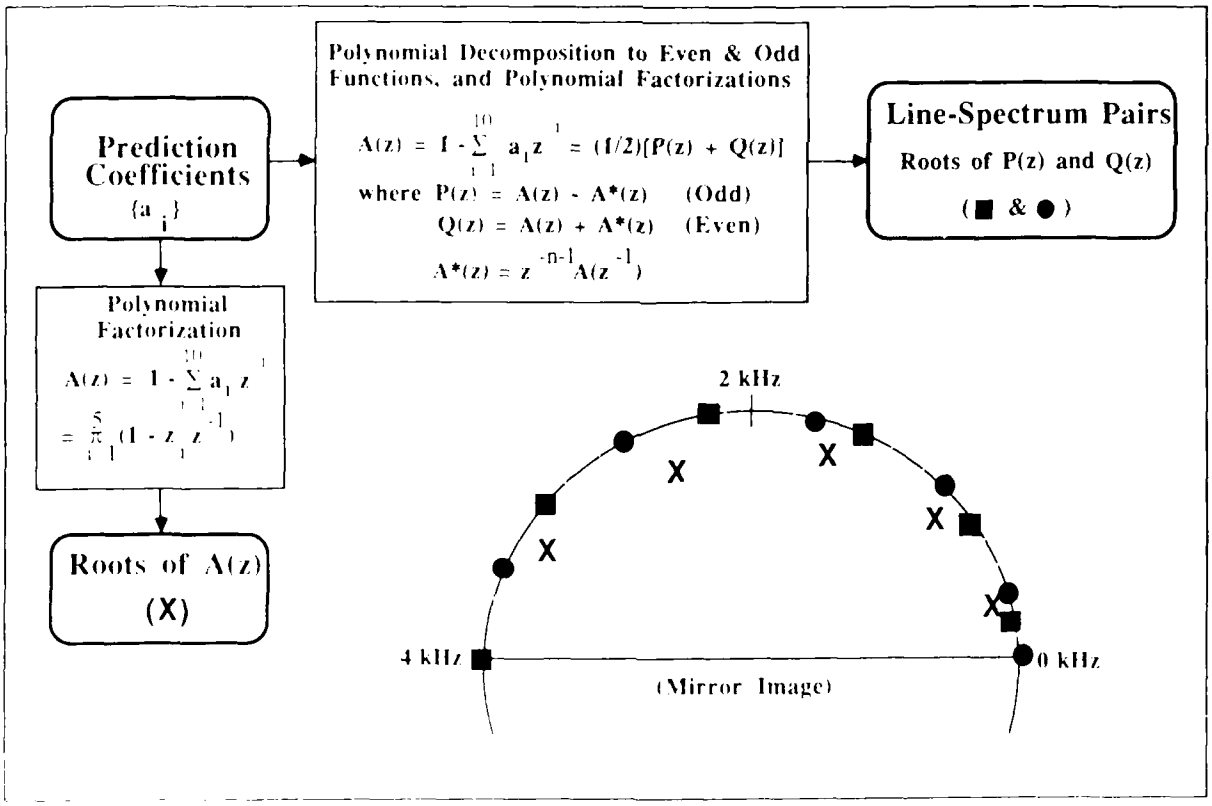


Fig. 5 — Decomposition of the roots of the LPC analysis filter $A(z)$. Since the LPC synthesis filter is the inverse of the LPC analysis filter, these roots represent the speech spectral envelope. As noted, each of the roots of the LPC analysis filter $A(z)$ located inside the unit circle (indicated by X), may be decomposed to two roots along the unit circle. One root belongs to $P(z)$, indicated by ■ and the other belongs to $Q(z)$, indicated by ●. Through this decomposition process, two extraneous roots (at $z = 1$ and $z = -1$) are generated. These need not be encoded because they are time-invariant.

To alleviate computational difficulties in searching for roots in a two-dimensional space, the LPC analysis filter may be decomposed to a sum of two filters in which each filter has roots along the unit circle of the complex z-plane. This can be accomplished by taking a sum and difference between $A(z)$ and its conjugate function (i.e., the transfer function of the filter whose impulse response is a mirror image of $A(z)$:

$$P(z) = A(z) - z^{-(n+1)}A(z^{-1}), \tag{3}$$

and

$$Q(z) = A(z) + z^{-(n+1)}A(z^{-1}). \tag{4}$$

The LPC analysis filter, reconstructed by the sum of these two filters, is

$$A(z) = \frac{1}{2}[P(z) + Q(z)]. \tag{5}$$

Equation (5) is an equivalent representation of the LPC analysis filter $A(z)$ in which $P(z)$ and $Q(z)$ are component filters. We will encode the parameters of $P(z)$ and $Q(z)$.

The impulse response of $P(z)$ expressed by Eq. (3) is odd symmetric with respect to its midpoint. Thus one real root is at $z = 1$, and other roots are at $z = \text{EXP}(j2\pi f_k t_s)$ where f_k is a member of the kth LSP, t_s is the speech sampling time-interval, and $j = \sqrt{-1}$. Thus, $P(z)$ may be factored as:

$$\begin{aligned} P(z) &= (1 - z^{-1}) \prod_{k=1}^{n/2} (1 - \epsilon^{j2\pi f_k t_s} z^{-1})(1 - \epsilon^{-j2\pi f_k t_s} z^{-1}) \\ &= (1 - z^{-1}) \prod_{k=1}^{n/2} [1 - 2\cos(2\pi f_k t_s) z^{-1} + z^{-2}]. \end{aligned} \tag{6}$$

On the other hand, $Q(z)$ is even symmetric with respect to its midpoint. Thus, one real root is at $z = -1$, and other roots are at $z = \exp(j2\pi f'_k)$. Thus,

$$Q(z) = (1 + z^{-1}) \prod_{k=1}^{n/2} [1 - 2\cos(2\pi f'_k t_s) z^{-1} + z^{-2}], \tag{7}$$

where f'_k is the other member of the kth LSP. Both f_k and f'_k are yet to be determined when they are discussed.

PC-to-LSP Conversion

The conversion of PC to LSP consists, in essence, in finding the roots of $P(z)$ and $Q(z)$. Since the roots are along the unit circle, they may be found by searching for null frequencies of the amplitude spectra. Figure 6 shows the computational modules needed for estimating LSPs. Since a 256-point complex FFT is involved, the required computational load is not trivial. The current CPU used in the ANDVT would take approximately 3 ms (i.e., 13% of the frame). To implement this approach, a more powerful CPU would therefore be needed.

Spectral Analysis

Since the impulse responses of $P(z)$ and $Q(z)$ are real, their amplitude spectra may be obtained simultaneously through the use of a single complex fast Fourier transform (FFT) [16]. Initially, the impulse responses of $P(z)$ and $Q(z)$ are loaded in the real and imaginary input FFT buffers, respectively. Then the remaining 244 samples are zero-padded for Fourier transform. The real and imaginary parts of the output are descrambled to obtain the two sets of amplitude spectra [16]. A transform size of 512 provides a frequency resolution of 4000 Hz/256 = 15.625 Hz.

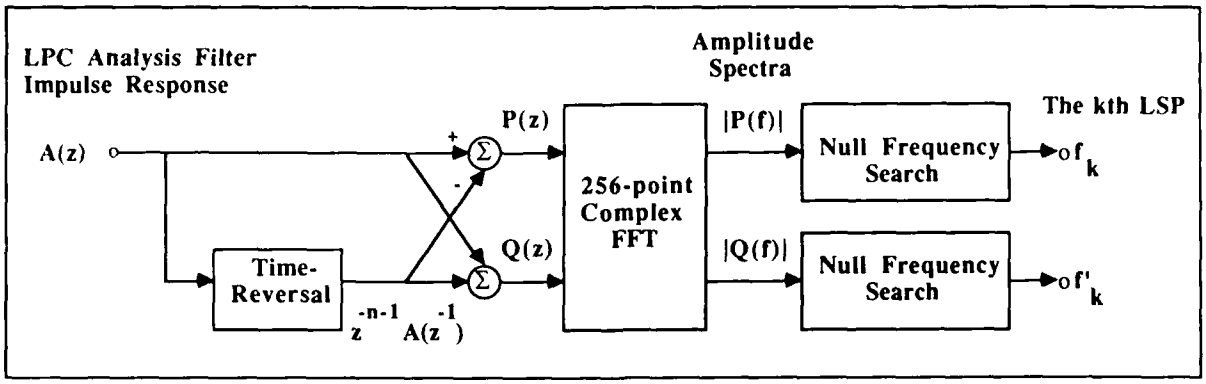


Fig. 6 — Computational modules required for estimating LSPs. For the 10th order LPC, the impulse response of the LPC analysis filter $A(z)$ is 11 samples. Thus both $P(z)$ and $Q(z)$ have 12 samples.

Search of Null Frequencies

Let the amplitude spectral components of either $P(z)$ or $Q(z)$ at frequency $f(j)$ be denoted by $y(j)$, $j=1, 2, \ldots, 256$. The line spectrum in the null frequency where the amplitude spectrum is at its local minimum. Thus three consecutive spectral points, $y(i-1)$, $y(i)$ and $y(i+1)$, have the following relationships near the null frequency $f(i)$:

$$y(i) < y(i-1), \qquad \text{for} \quad 2 \leqslant i \leqslant 255,$$

and

$$y(i) < y(i+1), \qquad \text{for} \quad 2 \leqslant i \leqslant 225. \tag{8}$$

Since the frequency resolution of the FFT is 15.625 Hz, the error in the estimated line spectrum is uniformly distributed between −7.8125 Hz and 7.8125 Hz. The estimated line spectrum, however, may be refined through a simple parabolic approximation based on the three consecutive spectral points (Fig. 7).

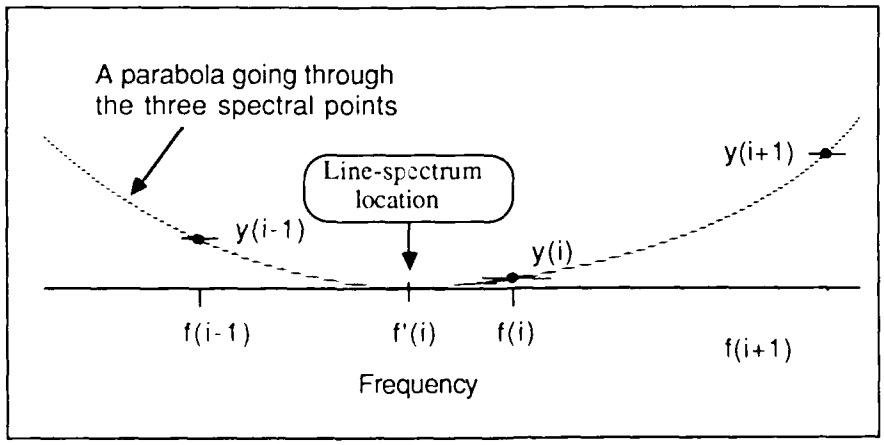


Fig. 7 — Refinement of the estimated line-spectral value through a parabolic fitting. If no frequency correction is made, $f(j)$ is the estimated line spectrum that would have an error somewhere between −7.8125 and 7.8125 Hz. If frequency correction is made, the estimated line spectrum is within a few Hz.

Substituting the three consecutive spectral points in the equation of a parabola, and after finding the solution for the frequency that makes the gradient of the parabola zero, we will have the refined line spectrum. Thus

$$f'(i) = f(i) + \frac{1}{2}\left\{\frac{y(i-1)-y(i+1)}{y(i-1)-2y(i)+y(i+1)}\right\}[f(i)-f(i-1)], \tag{9}$$

where $f'(i)$ is the refined line spectrum.

Figure 8 shows a typical picture of the LSP trajectories from actual speech samples. As noted, there are similarities between the trajectories of LSPs and speech resonant frequencies because both are frequency-domain parameters. Thus, an error in one line spectrum affects the synthesized speech spectrum only near that frequency. To exploit the listener's decreased sensitivity to frequency differences in the upper frequency region, we can quantize high-frequency LSPs more coarsely than low-frequency LSPs. This is a major advantage in this approach.

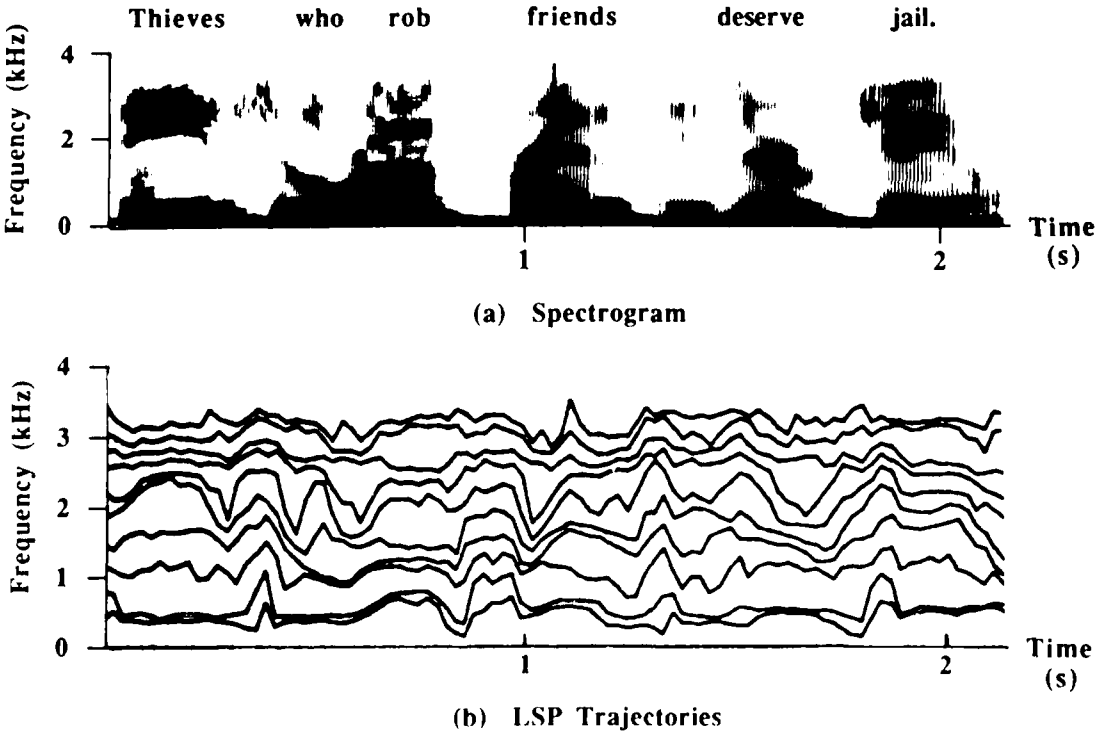


Fig. 8 — Typical LSP trajectories and spectrogram of the original speech. Since LSPs are located near the speech resonant frequencies, their trajectories are very similar

LSP-to-PC Conversion

The LSP-to-PC conversion is much more straightforward than the PC-to-LSP conversion. A set of LSPs can be converted to PCs by finding the solution for the coefficients of the polynomial that represent the transfer function of the LPC analysis filter, $A(z)$. Substituting Eqs. (6) and (7) into Eq. (5) gives

$$A(z) = \frac{1}{2}(1-z^{-1})\prod_{k=1}^{n/2}[1-2\cos(2\pi f_k t_s)z^{-1}+z^{-2}] + \frac{1}{2}(1+z^{-1})\prod_{k=1}^{n/2}[1-2\cos(2\pi f'_k t_s)z^{-1}+z^{-2}]. \tag{10}$$

When the product terms are multiplied out, the resultant polynomial is in the following form:

$$A(z) = 1 + \beta_1 z^{-1} + \beta_2 z^{-2} + \ldots + \beta_n z^{-n}. \quad (11)$$

Comparing term by term with Eq. (1) indicates that the ith prediction coefficient is $-\beta_i$ (where $1 \leqslant i \leqslant n$).

800-B/S VOICE ENCODER/DECODER

Bit Allocations

According to our experimentation, the most critical factor affecting speech intelligibility is the number of bits assigned to encode the *filter parameters.* Hence we encode both the *pitch period* and *speech amplitude* parameters as coarsely as the ear can tolerate. The remaining bits are allocated to encode the filter parameters.

(a) Pitch Period

The pitch period is encoded into five bits (12 steps/octave with a frequency range from 66.67 to 400 Hz). The pitch resolution is perceptually adequate so there will be no impression of a singing inflection, although the pitch is quantized to the chromatic equitempered scale. Since the pitch does not change too radically in normal conversation, it is transmitted only once every three frames.

(b) Amplitude Parameter

The amplitude parameter is the root-mean-square value of the speech waveform computed from each frame (i.e., every 22.5 ms). The amplitude parameter is quantized to 1 to 16 3 dB steps and transmitted once per each frame. In comparison with the 2400-b/s LPC, the resolution of the amplitude information is one bit less, but casual listening cannot detect the difference.

(c) Sync Bit

Since the pitch period is transmitted once every three frames, it is convenient to group three frames, and a sync bit is transmitted once for every three frames.

(d) Filter parameters

The remaining 12 bits are allocated to encode the filter parameters (Table 4) and are transmitted once per frame. As usual, the filter coefficients are encoded jointly (i.e., quantized vectorially through a pattern matching process). Such a quantization process results in efficient coding, because the reference filter parameters do not contain parameters from nonspeech sounds. In this approach, the given LSPs are compared with the stored LSP sets, and the index corresponding to the best matching LSP set is transmitted. The LSP encoder, therefore, has two functional modules: LSP template collection and template matching. The LSP quantization process is discussed separately from pages 13 through 17.

Table 4 — Bit Allocation Per Three Frames
for 800-b/s Voice Processor

Synchronization	1
Pitch Period	5
Amplitude Information	4+4+4 = 12
Filter Parameters (with voicing)	12+12+12 = 36
Total	54 bits

LSP Template Collection

Since 12 bits are allowed to encode the filter parameters, we use 4096 templates or patterns for the LSPs. Of the 4096 LSP templates, 3840 are for voiced speech and 256 for unvoiced speech. These figures are based on our experimentation with an 800-b/s voice processor that quantized reflection coefficients vectorially [14]. According to our subsequent experimentation with LSPs as filter parameters, we have no reason to change these figures.

Ideally, each LSP template produces the sound that is just noticeably different from the closest template. Because the human ear is insensitive to small differences in the patterns, each LSP in a given template has an *allowable frequency tolerance* (Fig. 9) within which there is no perceptible sound change. When each member of the LSP set falls inside the respective frequency tolerance of a reference LSP set, then the two sets are treated as equal.

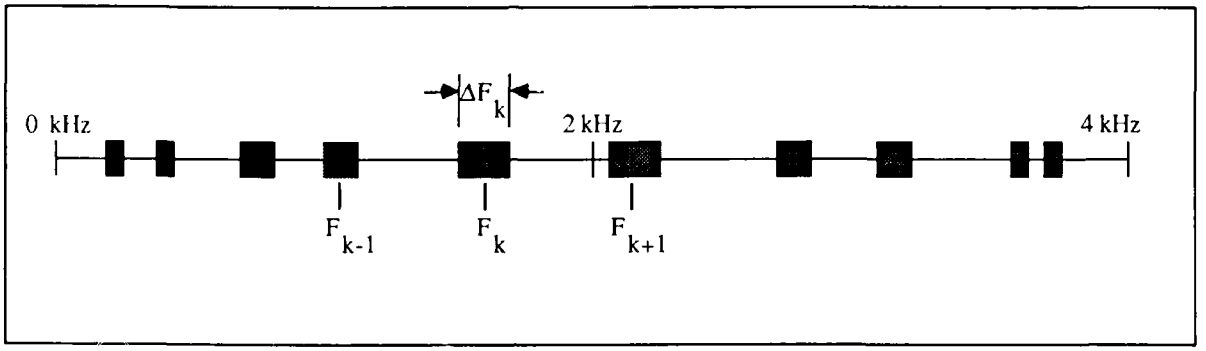


Fig. 9 — Frequency tolerance around each line spectrum. When each line spectrum is disturbed within its tolerance, the synthesized speech sounds no different. F_k is the kth line spectrum arranged in ascending order $F_1 < F_2 < \ldots < F_k < \ldots < F_{10}$.

During template collection, we initially store the first LSP set as a reference template. Subsequently, we compare each new LSP set with all the stored reference LSP templates. If the new LSP set falls outside the allowable frequency tolerance for every reference LSP, then the new LSP set becomes another reference LSP template (Fig. 10). In this investigation we used LSP templates collected from the voice of 54 males and 12 female speakers uttering five sentences each. During the template collection, the number of LSP sets that fell into each template was counted. At the end, the templates representing the fewest sets were eliminated to reduce the total number of templates to 4096.

Magnitude of LSP Frequency Tolerance

To utilize the 4096 LSP templates best, we have exploited both the ear's insensitivity to frequency differences and the LSP's tolerance of spectral errors.

(a) Hearing Sensitivity to Frequency Differences

Because the ear cannot resolve differences at high frequencies as accurately as it does at low frequencies, we may quantize higher frequency LSPs more coarsely than lower ones without introducing audible speech degradation. It is well known that the amount of frequency variation that produces a just-noticeable difference is approximately linear from 0.1 to 1 kHz, and it increases logarithmically from 1 to 10 kHz [17]. We documented a similar relationship for speech-like sounds using a pitch excitation signal with one of the ten line spectra incrementally changed while all others remained equal spaces (i.e., a resonant-free condition) [2]. Figure 11 shows the resulting curve. We expect that the curve of actual speech sounds would be located somewhere between these two curves. Figure 11 indicates that the frequency difference allowable near 4 kHz can be twice as large as that near 0 Hz.

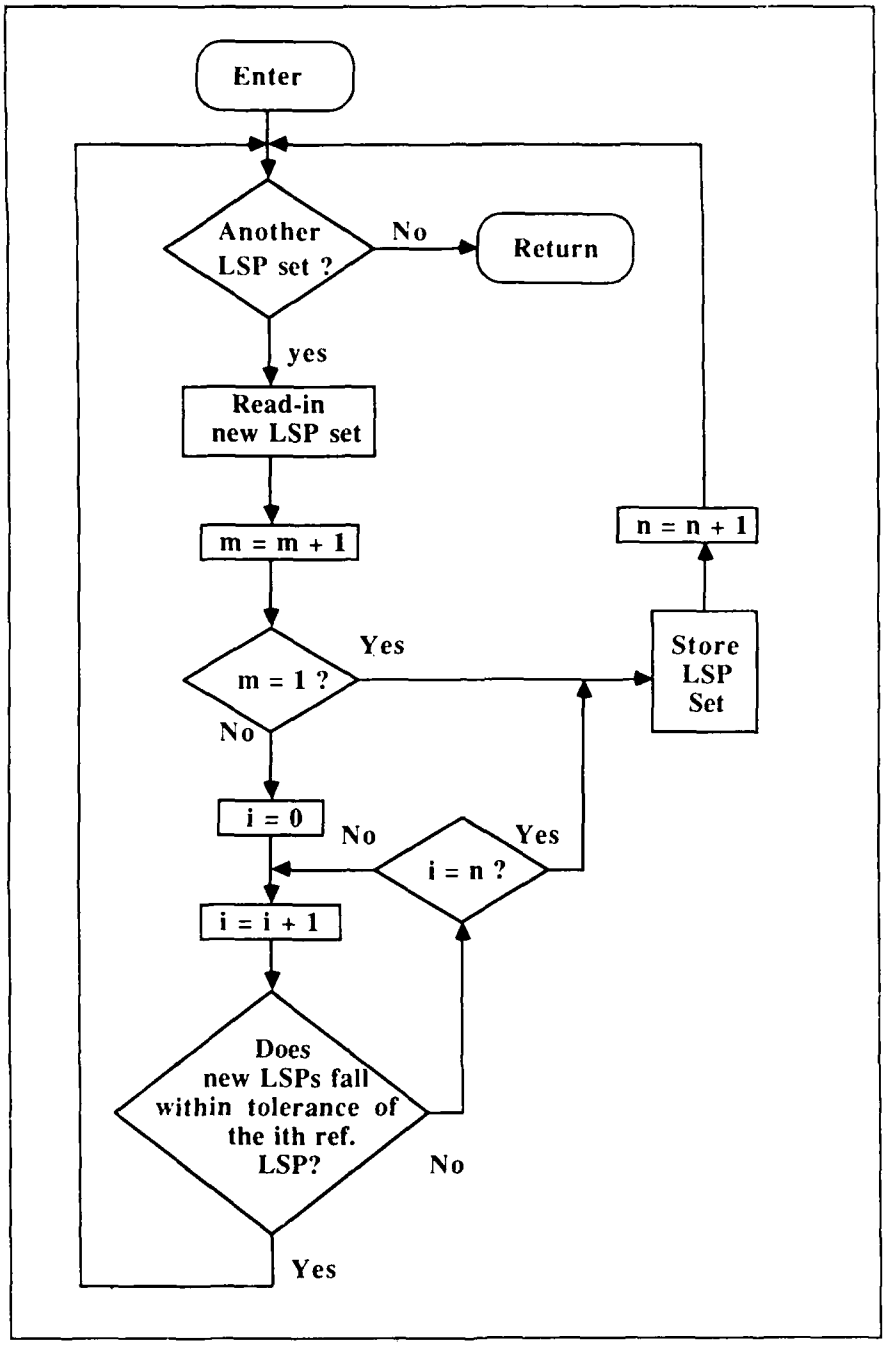


Fig. 10 — Flow diagram of LSP template collection process

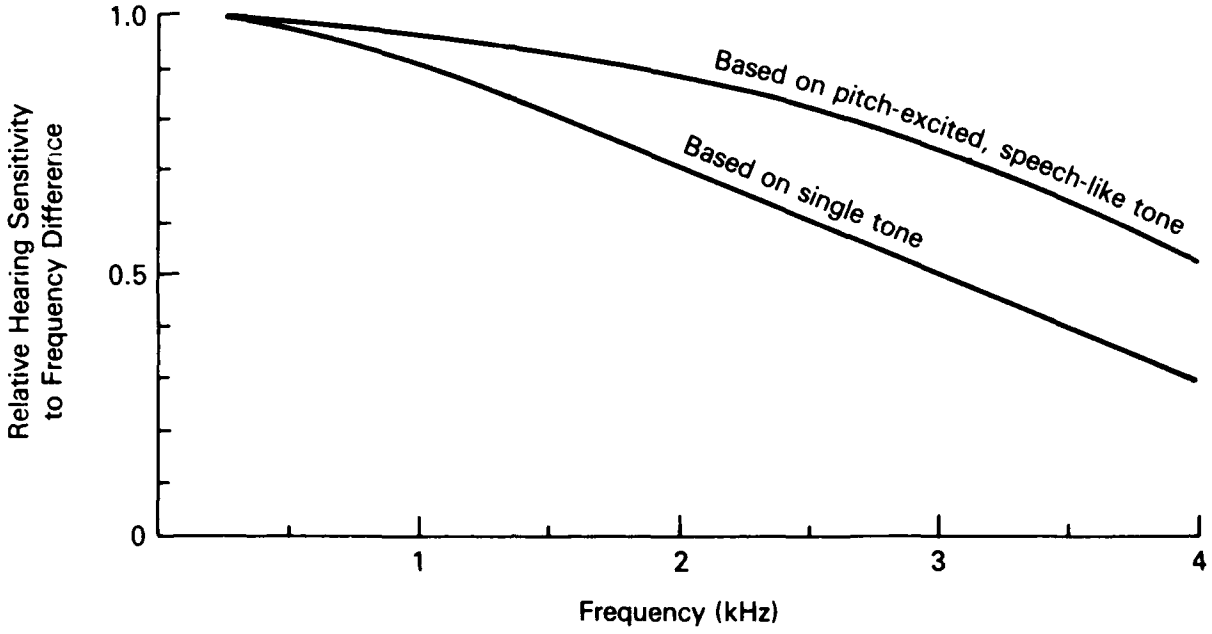


Fig. 11 — Relative hearing sensitivity to frequency differences

(b) Spectral Sensitivity of the LSP

When each line-spectrum is perturbed, there is a corresponding spectral error in $A(z)$. The spectral-error sensitivity is a factor relating error in each line-spectrum (in Hz) and the average spectral error of $A(z)$ (in dB). To derive such an expression from Eq. (10), however, is untractable. Also, a cross-coupling of all line-spectrum errors into the overall spectral error makes the use of such an expression impractical. Therefore, we derived numerically a relationship that relates the average spectral error of $A(z)$ to all of the line-spectrum errors (hence, including the effect of cross-couplings) from various speech samples. There is no approximation in computing the average spectral error of $A(z)$ from given line-spectrum errors. However, we imposed the condition that each line spectrum must have an error proportional to the frequency separation to its closest neighbor indicated in Fig. 9. Figure 12 is a resultant scatter plot. In our judgment, a 2 dB average spectral error is as big an error as we can tolerate. Thus the allowable frequency tolerance of each line spectrum as obtained from Fig. 12 is approximately 20% of the frequency separation to its closest neighbor.

(c) Allowable Frequency Tolerance

Combining the effect of the hearing sensitivity to the frequency difference (Fig. 11) and the spectral sensitivity of the LSP (Fig. 12), we have an allowable frequency tolerance for each LSP (see Fig. 13).

As shown in Fig. 13, the allowable frequency tolerance is approximately 20, 30, and 40% of the frequency separation to the closest neighbor for line spectra located below 1 kHz, between 1 and 2 kHz, and above 2 kHz, respectively. To verify this, we listened to many synthesized speech samples while perturbing each line spectrum by a given amount. Indeed, we began to notice some speech quality degradation when the perturbation exceeded the above-mentioned tolerance.

Template Matching

The LSPs in each frame are compared with all of the LSP templates, and the index corresponding to the closest match is transmitted. The template matching process (Fig. 14) computes the distance to each template, while taking into account the spectral-error sensitivity and the hearing sensitivity.

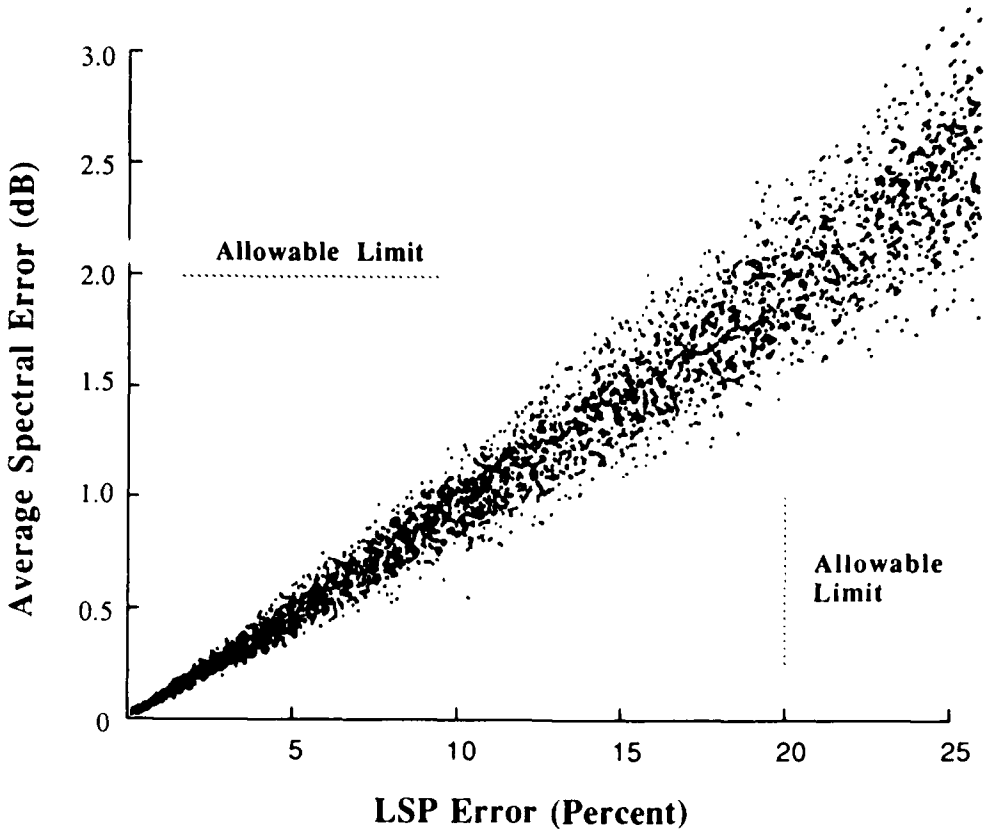


Fig. 12 — Scatter plot of average spectral error caused by the error in each line spectrum

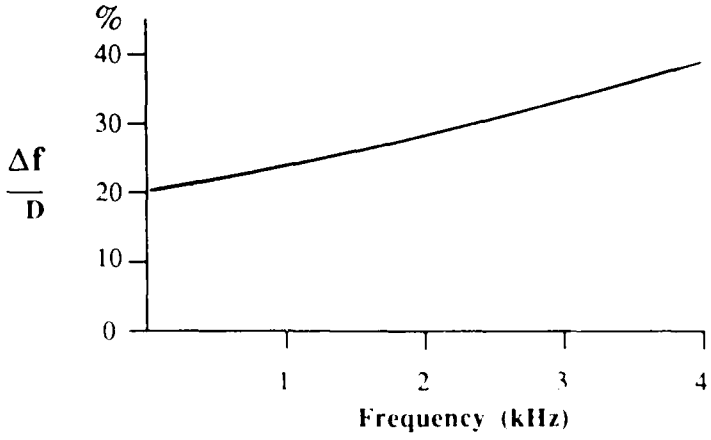


Fig. 13 — Allowable frequency tolerance of each line spectrum based on both the ear's sensitivity to frequency differences and the spectral sensitivity of the LSP for a 2 dB average error. Neither f nor D has an LSP index because this tolerance is applicable to any line spectrum

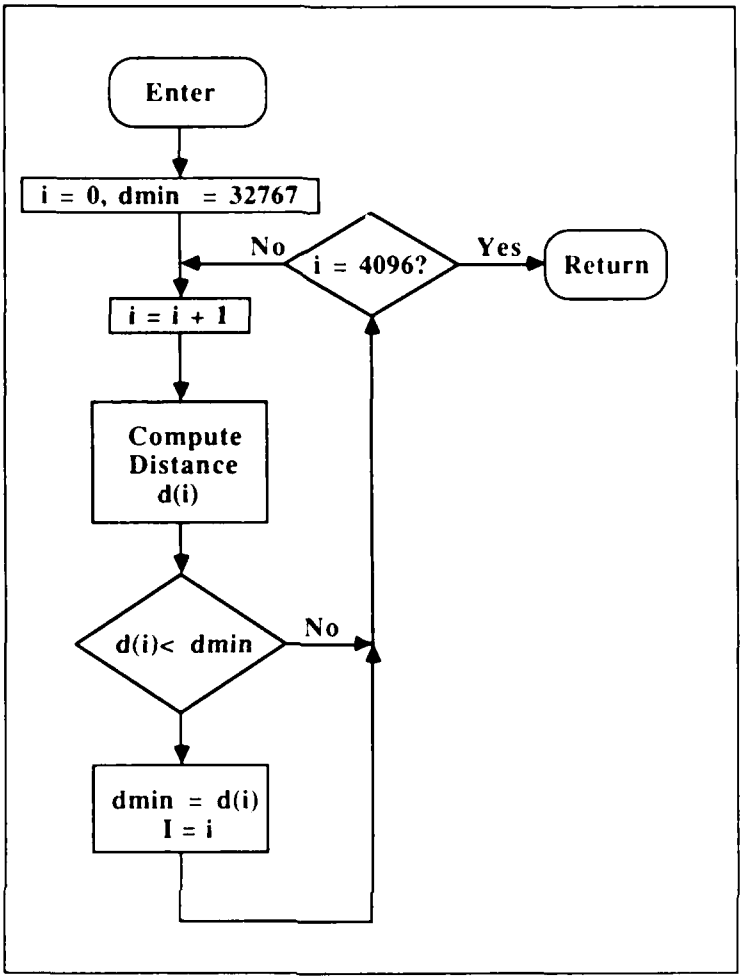


Fig. 14 — Flow diagram of LSP matching process

Though an exhaustive search of 4096 templates would appear to be a problem, our 800 b/s voice encoder used earlier was able to perform the task in real time with templates of ten reflection coefficients [15]. Searching 4096 LSP templates should be no problem by using the current technology.

Speech Intelligibility vs Bit-Error Rate

Figure 15 shows the intelligibility of the 800-b/s voice encoder that is discussed in this section under conditions of various bit-error rates. Although bit-errors may not be random in real environments, the use of random bit-errors for testing purpose is helpful for determining the strengths and weaknesses of the voice processor under investigation. Also, we have similar data from tests of other voice processors that allow us to compare and evaluate.

The *rate* of intelligibility loss caused by the random bit-errors is nearly identical among different voice processors operating at different rates as we have seen. Figure 15 shows a similar trend. Thus, intelligibility in the error-free condition can be used to predict the performance under bit errors. For this particular 800-b/s voice encoder, the bit-error rate should be less than approximately 2% to ensure adequate intelligibility.

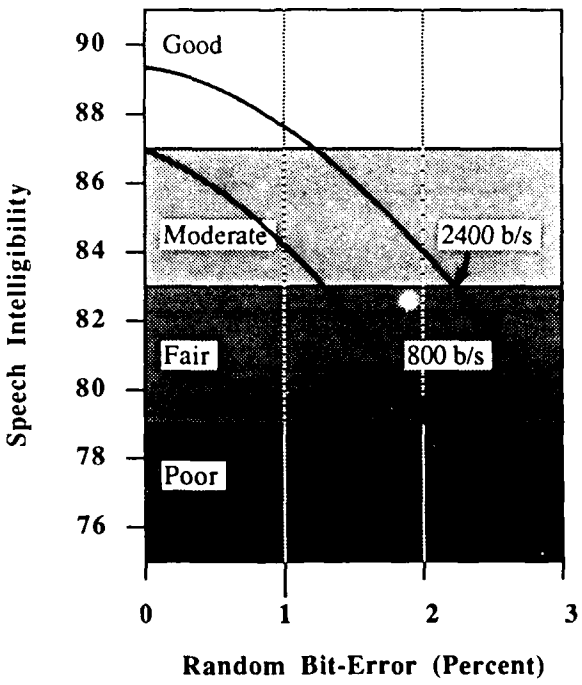


Fig. 15 — Speech intelligibility (DRT score) vs random bit-error rate. The 800 b/s voice encoder will be usable if the corrected bit-error rate is limited to 2%.

ERROR PROTECTION

For this report, we have investigated the potential advantages of providing error protection in an HF modem to all of the data bits. This is in contrast to the present 2400-b/s ANDVT TACTERM (CV-3592), that applies error protection to only the 24 most sensitive bits in each 54-bit LPC frame (see Table 2). The remaining 30 bits are transmitted without any error protection. To maintain as much common design between the present 2400-b/s system and the 800-b/s system presented in this report, the modulation was restricted to a four-phase differential phase shift key (DPSK) frequency division multiplex with a frame rate identical to the 2400-b/s LPC (i.e., 44.444 frames/s) and with tones spaced 56.25 Hz apart.

Simulated HF Channel and Signal Designs

An independent Rayleigh fading channel was used to compare the performance of the 800-b/s system with the ANDVT TACTERM. The comparisons were made by using four different signal designs for the 800-b/s system. Their characteristics were:

- 800 b/s transmission rate on 9 tones with no coding or diversity;
- 1600 b/s transmission rate on 18 tones with dual diversity;
- 3200 b/s transmission rate on 36 tones with quadruple diversity;
- 3200 b/s transmission rate on 36 tones with 1/2 rate (24,12) Golay coding on all of the 18 information bits per frame and transmitted with dual diversity. We used soft decision decoding, identical to that used in the ANDVT TACTERM.

The independent Rayleigh fading channel is a textbook channel. It is a transmission channel that exhibits fading with a Rayleigh amplitude distribution [18] with additive Gaussian noise that is independent on each of the modem subchannels. That is, there is no correlation in the fading on the different modem tones, which is usually not true on a real HF channel. An independent Rayleigh fading channel is excellent for determining the potential advantages of diversity combining and for coding that cannot be interleaved over many frames to randomize bursts.

We used Monte-Carlo simulation [19] for demonstration. It consists of the repetitive generation and demodulation of the received signal and its reference signal for each of the N tones in a modem frame. In a time-differential PSK system, the reference signal is the signal detected during the previous frame. It may be represented by two expressions that describe the in-phase and quadrature-phase components that would be obtained by correlating the received signal against a locally generated signal. The received signal during the present frame may be represented by two similar expressions. For the Rayleigh fading channel, the four expressions are:

$$\text{In-phase, reference:} \quad V_1 = VR_1 \cos(\phi_{R_1} + \phi_1) + X_1 \tag{12}$$

$$\text{Quadrature, reference:} \quad V_2 = VR_1 \sin(\phi_{R_1} + \phi_1) + Y_1 \tag{13}$$

$$\text{In-phase, signal:} \quad V_3 = VR_2 \cos(\phi_{R_2} + \phi_2 + \phi_D) + X_2 \tag{14}$$

$$\text{Quadrature, signal:} \quad V_4 = VR_2 \sin(\phi_{R_2} + \phi_2 + \phi_D) + Y_2 \tag{15}$$

where ϕ_1 is the reference phase shift (that was set equal to zero in this simulation) and ϕ_2 is the phase shift encoded in the transmitted signal. It was made equal to $\pi/4$ for all data symbols, which was equated to transmitting an all zero word. ϕ_D is phase shift caused by the doppler (that was also set to zero in the present simulation). The X and Y values were the in-phase and quadrature components of the additive Gaussian noise. The quantity V is a variable that controls the signal energy to noise density ratio expressed as the energy per tone to noise density ratio E_t/N_0):

$$\left[\frac{E_t}{N_0}\right] = 10 \log\left[\frac{V^2}{2}\right] \qquad dB \tag{16}$$

The quantity E_t/N_0 is related to the total signal energy to noise density P/N_0 by

$$\left[\frac{P}{N_0}\right] = \left[\frac{E_t}{N_0}\right] + 10 \log\left[\frac{N}{T}\right] \qquad dB \tag{17}$$

where N is the number of tones transmitted and T is the integration period that is the reciprocal of the tone spacing.

For a Rayleigh fading channel in which the fade rate is slow compared to the modem signaling rate (i.e., frame rate), then,

$$R_1 = R_2 \tag{18}$$

and

$$\phi_{R_1} = \phi_{R_2}, \tag{19}$$

which represents instantaneous samples of the channel-induced amplitude and phase variations on the received signal and its reference. In the simulation, each value was obtained by converting a random variable with uniform distribution to a random variable with a Gaussian distribution, and then converting that to a variable with a Rayleigh distribution [20].

Two Gaussianly distributed random variables with zero mean and a unit variance, X and Y, were obtained by

$$X = -2[\ln(A)]\cos(2\pi B) \tag{20}$$

$$Y = 2[\ln(A)]\sin(2\pi B) \tag{21}$$

where A and B were variables randomly selected from a set with uniform distribution (0,1). Likewise, a sample from a Rayleigh amplitude distribution with a unit variance was obtained as

$$R = 0.707\sqrt{X^2 + Y^2} \tag{22}$$

with a phase angle

$$\phi = \tan^{-1}\left(\frac{Y}{X}\right). \tag{23}$$

Demodulation

The demodulation of the received signal was performed to recover an estimate of the transmitted information. The in-phase and quadrature components of the phase change of the received signal relative to the reference signal were

$$I = V_1V_3 + V_2V_4 \tag{24}$$

$$Q = V_1V_4 - V_2V_3. \tag{25}$$

For Grey coded four-phase DPSK with two bits of information transmitted on each tone, the sign of I represented one information bit and the sign of Q represented the second bit of information. When diversity combining was performed, the values of I and the values of Q were added to those of a previous detection. Thus, for diversity

$$I_{div} = I_1 + I_2 \tag{26}$$

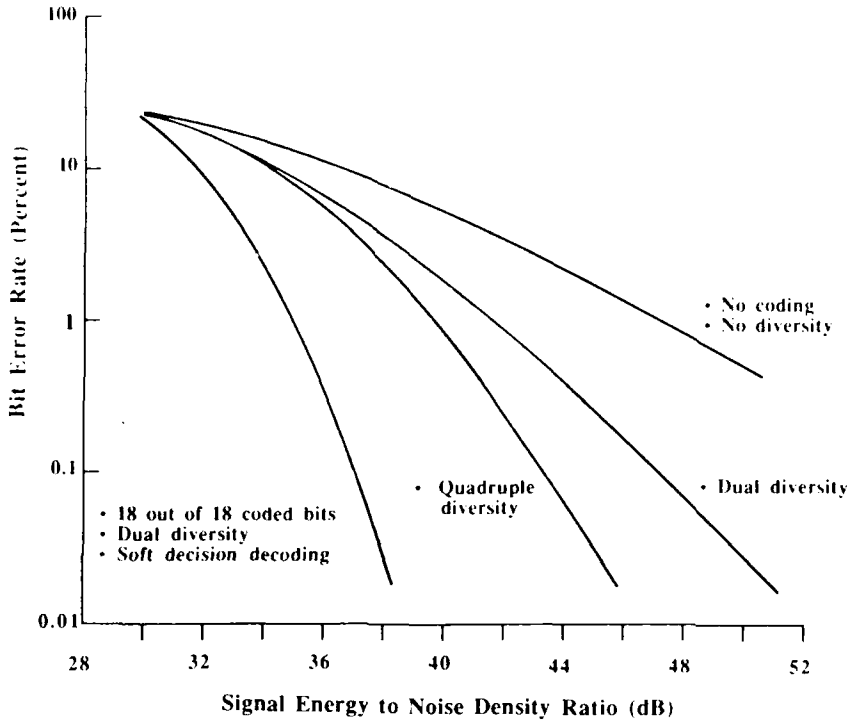
$$Q_{div} = Q_1 + Q_2 \tag{27}$$

and the detection of the received data is made on the signs of I_{div} and Q_{div}.

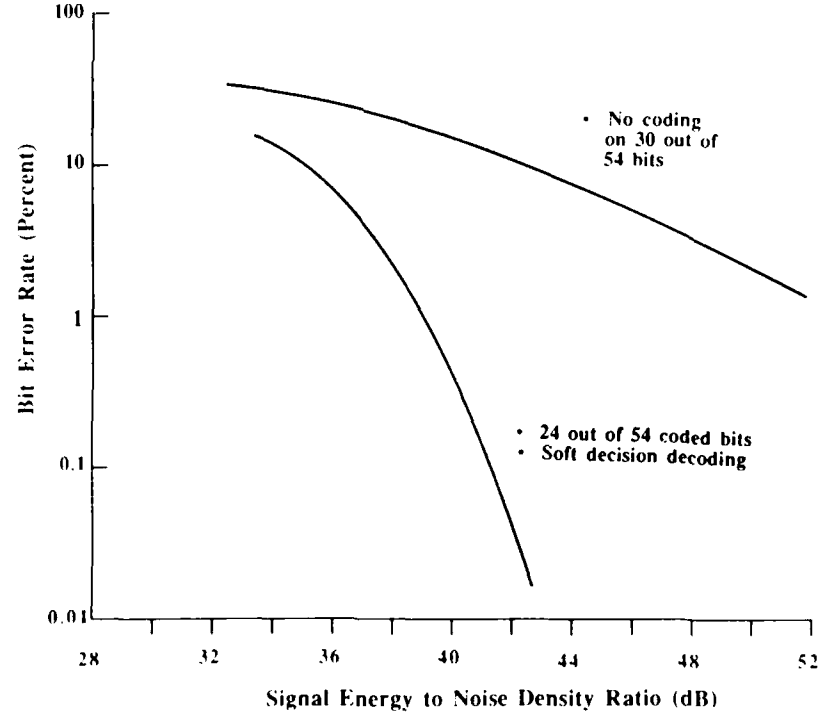
The soft decision decoding algorithm [21] was based on making up to 16 separate trials at decoding each received code word of 24 bits, using all permutations of the four bits with the lowest confidence. The best estimate of the correct data was obtained by selecting the decoding that indicated the errors were on the combination of bits with the lowest overall confidence. In this simulation of a coded system, separate code words were assigned to the in-phase and quadrature components, thus reducing the possibility of multiple errors in a code word when one signal was severely faded. That is similar to the code assignments used in ANDVT TACTERM.

Modem Performance

Figure 16 shows the average bit rates of the 800- and 2400-b/s designs. They are plotted according to the total signal energy to noise density ratio (P/N_0). Figure 16(a) clearly shows the advantage of using dual diversity to provide an initial improvement of 5 dB at a bit error rate of 1% followed by coding with soft decision decoding to give a total improvement of 12 dB at 1% error rate over the straight 800 b/s design. In Fig. 16(b) the 2400 b/s design shows a similar improvement between uncoded and coded voice data.



(a) 800-b/s System



(b) 2400-b/s System

Fig. 16 — Performance of 800 and 2400-b/s systems with different four-phase DPSK signal designs in a slow independent Rayleigh fading channel

Figure 17 shows a comparison between the 800-b/s design with coding and dual diversity (Fig. 16a) and 2400-b/s design with coding (Fig. 16b). As noted, the 800-b/s design has a 4-dB advantage at a bit error rate of 1 to 2%. This is a 4-dB advantage over the coded portion of the 2400-b/s system. The other 30 bits of the 2400-b/s system are transmitted uncoded and they contribute the intelligibility only under very low bit error conditions. At high bit error rate the 30 uncoded bits are a liability.

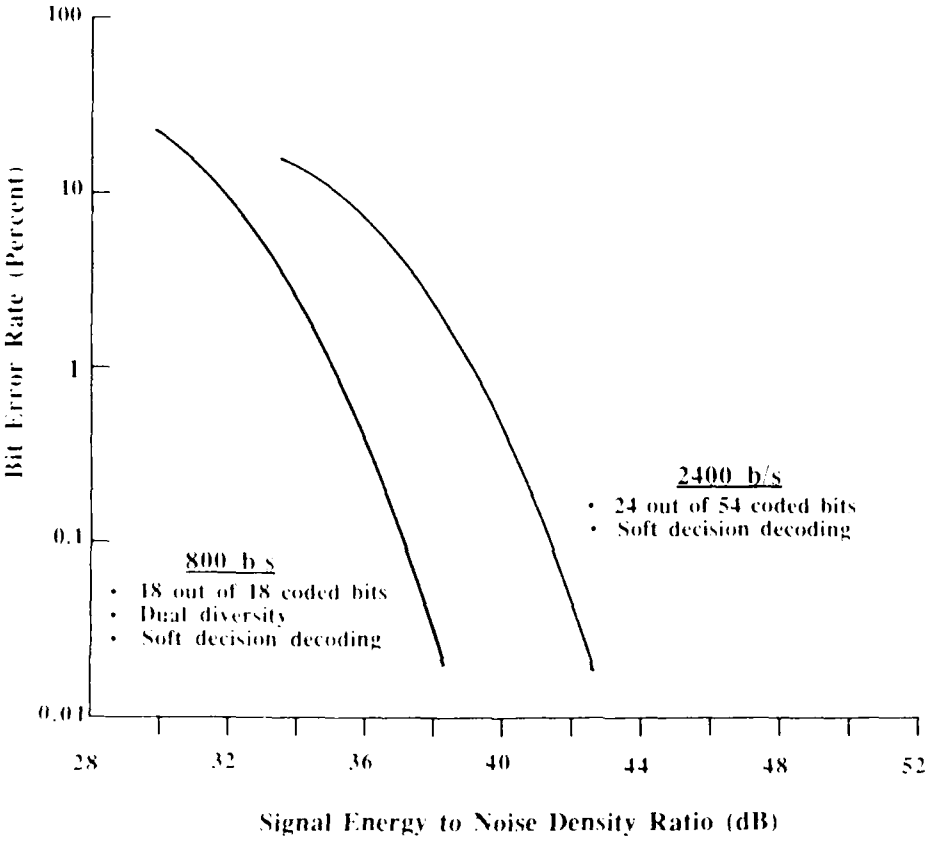


Fig. 17 — Comparison between 800-b/s design with coding and dual diversity and 2400-b/s design. The 800-b/s design has an advantage of nearly 4 dB over the 2400-b/s design. This advantage is equivalent to receiving two and a half times more signal power.

Figure 18 shows speech intelligibility in terms of the signal energy to noise density ratio. This figure is obtained by juxtaposing Fig. 17 (bit-error rate vs P/N_0) and Fig. 15 (intelligibility vs bit-error rate). It is significant that speech intelligibility degrades from *good* to *poor* with a 2 dB reduction in P/N_0. When the 2400-b/s system operates near the knee of the performance curve, the use of the 800-b/s system is much preferred. This report shows that the usable range of P/N_0 can be extended by nearly 4 dB.

CONCLUSIONS

This report discusses the result of our efforts to improve voice communication in the presence of bit errors. In particular, this improvement is designed for tactical communicators who use primarily narrowband channels and operate in congested platforms in close proximity to hostile forces. We have generated a more robust voice coding algorithm that can be integrated into the existing narrowband

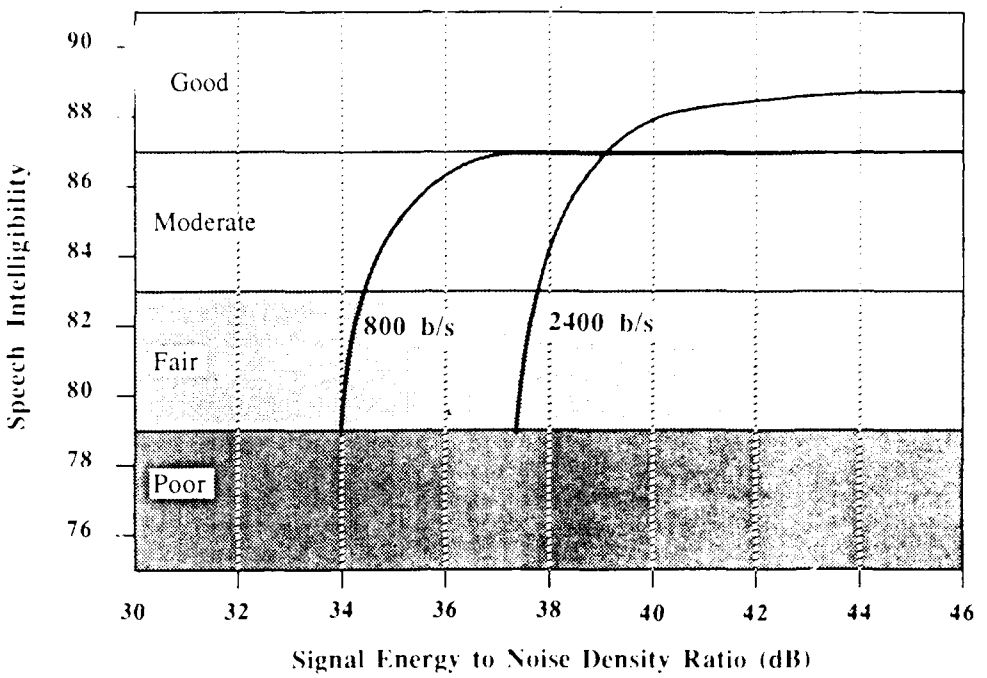


Fig. 18 — Speech intelligibility vs total signal energy to noise density ratio. Note that when the 2400-b/s system operates near the knee of the performance curve, the use of the 800-b/s design is preferred. In terms of speech intelligibility, the 800-b/s design has an advantage of 3.5 dB over the 2400-b/s design.

voice terminal so that the communicator can select either the DoD-standard 2400-b/s LPC that is interoperable with all narrowband users or the optional mode presented in this report.

Improved error-resistant performance is obtained by removing speech redundancies to lower the data rate from 2400 to 800 b/s, and then introducing other redundancies in the form of frequency diversity and coding to provide error protection. To simplify the implementation, we have maintained the basic feature of the ANDVT in speech processing, error protection, and modem designs.

We chose a slow independent Rayleigh fading channel to make a performance comparison between the 2400- and 800-b/s systems. The most significant conclusions follow.

- The error rate for the 800-b/s system is one order of magnitude less than that for the 2400-b/s system for a wide range of signal energy to noise density ratios (Fig. 17).
- For an error rate of 2% or less, the 800-b/s system has a 4 dB advantage in the signal energy to noise density ratio over the 2400-b/s system (Fig. 17).
- When the 2400-b/s system provides poor speech quality, the 800-b/s system provides better speech quality even when the signal energy to noise density is 3.5 dB less (Fig. 18). In other words, the 800-b/s system behaves like the 2400-b/s system operating under 2.5 times more signal power.

This report represents an initial attempt to provide a more robust performance for narrowband users; it is intended to create interest among DoD policy makers, program sponsors, and system designers. We think this approach is worthy of a continued investigation.

RECOMMENDATIONS

Prior to committing prototype implementation, we recommend the following tasks:

- The 800-b/s voice processor should be programmed to run in real time, not only to allow performance of additional tests but to gain experience in generating the real-time software.
- The overall performance should be further evaluated by using other forms of channel disturbances.
- Efforts should be continued to develop a voice processor capable of generating intelligible speech at lower bit rates.

ACKNOWLEDGMENTS

This work is funded by the Office of Naval Research and NAVSPAWARSYSCOM. The authors thank Drs. J. Davis of NRL and R. Martin and R. Allen of NAVSPAWARSYSCOM for their support. The authors extend thanks to Larry Fransen and Stephanie Everett for their constructive suggestions for this report.

REFERENCES

1. "Analog to Digital Conversion of Voice by 2,400 Bits/Second Linear Predictive Coding," *Federal Standard 1015*, GSA, 7th and D Streets, S.W., Washington, DC 20407, Nov. 28, 1984.

2. G.S. Kang and L.J. Fransen, "Low-Bit Rate Speech Encoders Based on Line-Spectrum Frequencies (LSFs)," NRL Report 8857, Jan. 1985.

3. G.S. Kang and L.J. Fransen, "Application of Line-spectrum Pairs to Low-Bit-Rate Speech Encoders," IEEE ICASSP Conference Record, p. 244-247, 1985.

4. G.S. Kang and S.S. Everett, "Improvement of the Narrowband Linear Predictive Coder," Part 1—Analysis Improvements, NRL Report 8645, Dec. 1982; Part 2—Synthesis Improvements, NRL Report 8799, June 1984.

5. G.S. Kang and S.S. Everett, "Improvement of the Excitation Source in the Narrow-band Linear Prediction Vocoder," *IEEE Trans. Acoustics, Speech and Signal Proc.* **ASSP-33**, 377-386, 1985.

6. G.S. Kang, "Narrowband Integrated Voice/Data System Based on the 2400-b/s LPC," NRL Report 8942, Dec. 1985.

7. G.S. Kang and L. Fransen, "Experimentation with an Adaptive Noise-Cancellation Filter," submitted for publication in *IEEE Trans. Circuits and Systems,* Aug. 1986.

8. W.M. Jewett and R. Cole, Jr., "Modulation and Coding Study for the Advanced Narrowband Digital Voice Terminal," NRL Memorandum Report 3811, Aug. 1978.

9. W.M. Jewett and R. Cole, Jr., "Non-Real Time Stress Tests of the ANDVT HF Modem," NRL Memorandum Report 4574, Aug. 10, 1981.

10. G.S. Kang and D.C. Coulter, "600-Bits-Per-Second Voice Digitizer," NRL Memorandum Report 8043, Nov. 1976.

11. G.S. Kang and D.C. Coulter, "600 bps Voice Digitizer," 1976 IEEE ICASSP Conference Record, p. 91-94.

12. B.H. Juang, D.Y. Wong, and A.H. Gray, Jr., "Distortion Performance of Vector Quantization for LPC Voice Coding," *IEEE Trans. Acoustics, Speech and Signal Proc.* **ASSP-30** (2) (1982).

13. T.E. Carter, D.M. Dlugos, and D.C. LeDoux, "An 800 BPS Real-Time Voice Coding System Based on Efficient Encoding Techniques," IEEE ICASSP Conference Record, pp. 602-605, 1982.

14. L.J. Fransen, "2400- to 800-b/s LPC Rate Converter," NRL Report 8716, June 1983.

15. L.J. Fransen, "Technical Evaluation of Low Data Rate Experimental Terminal (LDRET)," Internal Technical Memodandum, June 5, 1985.

16. E.O. Brigham, *The Fast Fourier Transform*, (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974).

17. P. Ladefoged, *Elements of Acoustic Phonetics*, (The University of Chicago Press, Chicago and London, 1974).

18. A.D. Whalen, *Detection of Signals in Noise* (Academic Press, Inc., N.Y., 1971).

19. P. Beckman, *Probability in Communication Engineering* (Harcourt, Brace & World, N.Y., 1967).

20. G.M. Dillard, "Generating Random Numbers Having Probability Distributions Occurring in Signal Detection Problems," *IEEE Trans.* **IT-13** (4), 616-617 (1967).

21. D. Chase, "A Class of Algorithms for Decoding Block Codes with Channel Measurement Information," *IEEE Trans.* **IT-18** (1), 170 (1972).

END

4-87

DTIC